## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (http://bmjopen.bmj.com/site/about/resources/checklist.pdf) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

## **ARTICLE DETAILS**

TITLE (PROVISIONAL)	Clinical decisions presented to patients in hospital encounters: a
	cross-sectional study using a novel taxonomy
AUTHORS	Ofstad, Eirik; Frich, Jan; Schei, Edvin; Frankel, Richard; Šaltytė Benth, Jūratė; Gulbrandsen, Pål

## VERSION 1 – REVIEW

REVIEWER	Anne Stiggelbout
	Leiden University Medical Center, Netherlands
REVIEW RETURNED	23-Jun-2017
GENERAL COMMENTS	I am not a native speaker, and neverteless found some phrases (e.g. amount of decisions instead of number, less instead of fewer, a statement that is said) that do not sound right, so I suggest having a native speaker check the english. I miss some comma's and articles as well, that would increase readability. And sentences should not start with a number (a percentage, e.g. P9, L 14, P10, L 52 and 54)
	I commend the authors for their extensive work on classifying decisions, particularly with these large numbers. Most of my queries/questions, relate to their definition of a decision and their classification, which is somewhat difficult because it has already been published by BMJOpen. So most of my comments are textual, of necessity.
	Even though the classification is out in the open, I have been googling, and depending on the website accessed, decisions and judgments are treated as almost the same or different. The authors build on Braddock: "a decision is a verbal statement committing to a particular course of action", and have simply added: "and/or statement concerning the patient's health that carries meaning and weight because it is said by a medical expert". This to me opens up the definition to any statement by a doctor concerning the patient's health (even false statements!). So then a judgment, a recommendation, anything related to the patient's health becomes a decision? Doctor:"mr Jones, you are obese" now is a decision? With which I personally disagree. The example given on blood pressure exemplifies this: "Your blood pressure is 180/100. This is high." So the first statement is a fact, the second a decision (according to the authors). I would consider this a judgment. "Given your age, this means: high." For me the decision is: "we need to do something about this" (or even better, "we have three options do nothing, change behaviour take pills" and which
	one is decided upon to me would be the decision).



<ol> <li>Since the readers should not need to go back to the 2016 paper, I would like to see a comment on the coding of the Preformed decisions. A limitation is that the coder cannot always no if the decision was preformed. E, in oncology, many decisions have been preformed in the Multidisciplinary Team Meetings, but this may not be stated as such to the patients. The doctor will simply say, "we". (See Engelhardt et al. EU2 2016, page 59, left column, second para). So I am convinced there will have been more preformed than counted. Please comment on this in your Discussion. This is stated on P12, L 34, for WR decisions, but holds for many specialties nowadays too.</li> <li>Textual:</li> <li>P3 Not all readers will see why the statistical analysis is a strength of this study, please explain.</li> <li>P4 L 11-12 Since Finestein's conclusion was drawn in 1994, I would change "has left" in line 11-12 to "had left"</li> <li>P4 L 16-21 the paper mostly speaks of decisions to do something, I suggest adding -e.g. in this paragraph- doing nothing also as an outcome of a decision. Also e.g. p5, L11-16.</li> <li>P4 L 21 I do not see why this is "but", the pressure is not necessarily related to who makes the decision. I suggest rephrasing this sentence.</li> <li>P4 L34-39, EBM for a long time now has also included patient preferences in addition to best available evidence!</li> <li>P3 L45: Research and implementation not only target single decisions, suggest adding Often. Further, I disagree with the phrasing "addresses decisions where medical evidence provides no clear guidance", to a naive reader this may sound as if SDM is only for cases where the doc does not know. But the clear guidance may be that there are two options that patients may weigh differently and thus the patient should be involved</li> <li>P6, L 40-45: the response should be in the Results section. And please provide the 58% response after the 59 who provided broad consent.</li> <li>P7 L 32: please explain how i</li></ol>	
<ul> <li>Textual:</li> <li>P3 Not all readers will see why the statistical analysis is a strength of this study, please explain.</li> <li>P4 L 11-12 Since Finestein's conclusion was drawn in 1994, I would change 'has left' in line 11-12 to 'had left'</li> <li>P4 L 16-21: the paper mostly speaks of decisions to do something, I suggest adding -e.g. in this paragraph- doing nothing also as an outcome of a decision. Also e.g. p5, L11-16.</li> <li>P4 L 21 I do not see why this is "but", the pressure is not necessarily related to who makes the decision. I suggest rephrasing this sentence.</li> <li>P4 L 34-39, EBM for a long time now has also included patient preferences in addition to best available evidence!</li> <li>P3 L45: Research and implementation not only target single decisions, suggest adding Often. Further, I disagree with the phrasing' addresses decisions where medical evidence provides no clear guidance", to a naïve reader this may sound as if SDM is only for cases where the doc does not know. But the clear guidance may be that there are two options that patients may weigh differently and thus the patient should be involved</li> <li>P6, L 32: please explain Broad consent at its first use.</li> <li>P6, L 40-45: the response should be in the Results section. And please provide the 58% response after the 59 who provided broad consent.</li> <li>P7 L 32: please explain how intra-rater reliability was assessed. Did EHO code x video's twice, and how long was the period between the two?</li> <li>7 L 43=48: are "participants' patients? And please check parentheses, after old patients; and physicians. I suggest disentangling patients and physicians. I suggest disentangling pa</li></ul>	8. Since the readers should not need to go back to the 2016 paper, I would like to see a comment on the coding of the Preformed decisions. A limitation is that the coder cannot always no if the decision was preformed. E.g. in oncology, many decisions have been preformed in the Multidisciplinary Team Meetings, but this may not be stated as such to the patients. The doctor will simply say: "we". (See Engelhardt et al. EJC 2016, page 59, left column, second para). So I am convinced there will have been more preformed than counted. Please comment on this in your Discussion. This is stated on P12, L 34, for WR decisions, but holds for many specialties nowadays too.
<ul> <li>P3 Not all readers will see why the statistical analysis is a strength of this study, please explain.</li> <li>P4 L 11-12 Since Finestein's conclusion was drawn in 1994, I would change 'has left' in line 11-12 to 'had left'</li> <li>P4 L 16-21: the paper mostly speaks of decisions to do something, I suggest adding –e.g. in this paragraph-doing nothing also as an outcome of a decision. Also e.g. p5, L11-16.</li> <li>P4 L 21 I do not see why this is "but", the pressure is not necessarily related to who makes the decision. I suggest rephrasing this sentence.</li> <li>P4 L 34-39, EBM for a long time now has also included patient preferences in addition to best available evidence!</li> <li>P3 L45: Research and implementation not only target single decisions, suggest adding Offen. Further, I disagree with the phrasing "addresses decisions where medical evidence provides no clear guidance", to a naïve reader this may sound as if SDM is only for cases where the doc does not know. But the clear guidance may be that there are two options that patients may weigh differently and thus the patient should be involved</li> <li>P6, L 40-45: the response should be in the Results section. And please provide the 58% response after the 59 who provided broad consent.</li> <li>P7 L 32: please explain how intra-rater reliability was assessed. Did EHO code x video's twice, and how long was the period between the two?</li> <li>7 L 43=48: are "participants' patients? And please check parentheses, after old patients, and physicians. I suggest disentangling patients and physicians, for clarity.</li> <li>P7 L 52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.</li> </ul>	Textual:
<ul> <li>P4 L 11-12 Since Finestein's conclusion was drawn in 1994, I would change "has left" in line 11-12 to "had left"</li> <li>P4 L 16-21: the paper mostly speaks of decisions to do something, I suggest adding —e.g. in this paragraph-doing nothing also as an outcome of a decision. Also e.g. p5, L11-16.</li> <li>P4 L 21 I do not see why this is "but", the pressure is not necessarily related to who makes the decision. I suggest rephrasing this sentence.</li> <li>P4 L 34-39, EBM for a long time now has also included patient preferences in addition to best available evidence!</li> <li>P3 L45: Research and implementation not only target single decisions, suggest adding Often. Further, I disagree with the phrasing "addresses decisions where medical evidence provides no clear guidance". to a naive reader this may sound as if SDM is only for cases where the doc does not know. But the clear guidance may be that there are two options that patients may weigh differently and thus the patient should be involved</li> <li>P6, L 32: please explain Broad consent at its first use.</li> <li>P6, L 40-45: the response should be in the Results section. And please provide the 58% response after the 59 who provided broad consent.</li> <li>P7 L 32: please explain how intra-rater reliability was assessed. Did EHO code x video's twice, and how long was the period between the two?</li> <li>7 L 43=48: are "participants' patients? And please check parentheses, after old patients, and physicians. I suggest disentangling patients and physicians. For clarity.</li> <li>P7 L 52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.</li> </ul>	P3 Not all readers will see why the statistical analysis is a strength of this study, please explain.
<ul> <li>P4 L 16-21: the paper mostly speaks of decisions to do something, I suggest adding -e.g. in this paragraph- doing nothing also as an outcome of a decision. Also e.g. p5, L11-16.</li> <li>P4 L 21 I do not see why this is "but", the pressure is not necessarily related to who makes the decision. I suggest rephrasing this sentence.</li> <li>P4 L 34-39, EBM for a long time now has also included patient preferences in addition to best available evidence!</li> <li>P3 L 45: Research and implementation not only target single decisions, suggest adding Often. Further, I disagree with the phrasing "addresses decisions where medical evidence provides no clear guidance", to a naïve reader this may sound as if SDM is only for cases where the doc does not know. But the clear guidance may be that there are two options that patients may weigh differently and thus the patient should be involved</li> <li>P6, L 40-45: the response should be in the Results section. And please provide the 58% response after the 59 who provided broad consent.</li> <li>P7 L 32: please explain how intra-rater reliability was assessed. Did EHO code x video's twice, and how long was the period between the two?</li> <li>7 L 43=48: are "participants' patients? And please check parentheses, after old patients, and physicians. I suggest disentangling patients and physicians, for clarity.</li> <li>P7 L 52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.</li> </ul>	P4 L 11-12 Since Finestein's conclusion was drawn in 1994, I would change "has left" in line 11-12 to "had left"
<ul> <li>P4 L 21 I do not see why this is "but", the pressure is not necessarily related to who makes the decision. I suggest rephrasing this sentence.</li> <li>P4 L34-39, EBM for a long time now has also included patient preferences in addition to best available evidence!</li> <li>P3 L45: Research and implementation not only target single decisions, suggest adding Often. Further, I disagree with the phrasing "addresses decisions where medical evidence provides no clear guidance", to a naïve reader this may sound as if SDM is only for cases where the doc does not know. But the clear guidance may be that there are two options that patients may weigh differently and thus the patient should be involved</li> <li>P6, L 40-45: the response should be in the Results section. And please provide the 58% response after the 59 who provided broad consent.</li> <li>P7 L 32: please explain how intra-rater reliability was assessed. Did EHO code x video's twice, and how long was the period between the two?</li> <li>7 L 43=48: are "participants' patients? And please check parentheses, after old patients, and physicians. I suggest disentangling patients and physicians. For clarity.</li> <li>P7 L 52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.</li> </ul>	P4 L 16-21: the paper mostly speaks of decisions to do something, I suggest adding –e.g. in this paragraph- doing nothing also as an outcome of a decision. Also e.g. p5, L11-16.
<ul> <li>P4 L34-39, EBM for a long time now has also included patient preferences in addition to best available evidence!</li> <li>P3 L45: Research and implementation not only target single decisions, suggest adding Often. Further, I disagree with the phrasing "addresses decisions where medical evidence provides no clear guidance", to a naïve reader this may sound as if SDM is only for cases where the doc does not know. But the clear guidance may be that there are two options that patients may weigh differently and thus the patient should be involved</li> <li>P6, L32: please explain Broad consent at its first use.</li> <li>P6, L 40-45: the response should be in the Results section. And please provide the 58% response after the 59 who provided broad consent.</li> <li>P7 L 32: please explain how intra-rater reliability was assessed. Did EHO code x video's twice, and how long was the period between the two?</li> <li>7 L 43=48: are "participants' patients? And please check parentheses, after old patients, and physicians. I suggest disentangling patients and physicians, for clarity.</li> <li>P7 L52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.</li> </ul>	P4 L 21 I do not see why this is "but", the pressure is not necessarily related to who makes the decision. I suggest rephrasing this sentence.
<ul> <li>P3 L45: Research and implementation not only target single decisions, suggest adding Often. Further, I disagree with the phrasing "addresses decisions where medical evidence provides no clear guidance", to a naïve reader this may sound as if SDM is only for cases where the doc does not know. But the clear guidance may be that there are two options that patients may weigh differently and thus the patient should be involved</li> <li>P6, L 40-45: the response should be in the Results section. And please provide the 58% response after the 59 who provided broad consent.</li> <li>P7 L 32: please explain how intra-rater reliability was assessed. Did EHO code x video's twice, and how long was the period between the two?</li> <li>7 L 43=48: are "participants' patients? And please check parentheses, after old patients, and physicians. I suggest disentangling patients and physicians, for clarity.</li> <li>P7 L 52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.</li> </ul>	P4 L34-39, EBM for a long time now has also included patient preferences in addition to best available evidence!
<ul> <li>P6, L32: please explain Broad consent at its first use.</li> <li>P6, L 40-45: the response should be in the Results section. And please provide the 58% response after the 59 who provided broad consent.</li> <li>P7 L 32: please explain how intra-rater reliability was assessed. Did EHO code x video's twice, and how long was the period between the two?</li> <li>7 L 43=48: are "participants' patients? And please check parentheses, after old patients, and physicians. I suggest disentangling patients and physicians, for clarity.</li> <li>P7 L52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.</li> </ul>	P3 L45: Research and implementation not only target single decisions, suggest adding Often. Further, I disagree with the phrasing "addresses decisions where medical evidence provides no clear guidance", to a naïve reader this may sound as if SDM is only for cases where the doc does not know. But the clear guidance may be that there are two options that patients may weigh differently and thus the patient should be involved
<ul> <li>P6, L 40-45: the response should be in the Results section. And please provide the 58% response after the 59 who provided broad consent.</li> <li>P7 L 32: please explain how intra-rater reliability was assessed. Did EHO code x video's twice, and how long was the period between the two?</li> <li>7 L 43=48: are "participants' patients? And please check parentheses, after old patients, and physicians. I suggest disentangling patients and physicians, for clarity.</li> <li>P7 L52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.</li> </ul>	P6, L32: please explain Broad consent at its first use.
<ul> <li>P7 L 32: please explain how intra-rater reliability was assessed. Did EHO code x video's twice, and how long was the period between the two?</li> <li>7 L 43=48: are "participants' patients? And please check parentheses, after old patients, and physicians. I suggest disentangling patients and physicians, for clarity.</li> <li>P7 L52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.</li> </ul>	P6, L 40-45: the response should be in the Results section. And please provide the 58% response after the 59 who provided broad consent.
<ul> <li>7 L 43=48: are "participants' patients? And please check parentheses, after old patients, and physicians. I suggest disentangling patients and physicians, for clarity.</li> <li>P7 L52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.</li> </ul>	P7 L 32: please explain how intra-rater reliability was assessed. Did EHO code x video's twice, and how long was the period between the two?
P7 L52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.	7 L 43=48: are "participants' patients? And please check parentheses, after old patients, and physicians. I suggest disentangling patients and physicians, for clarity.
	P7 L52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.

P8, L 38-39: We analysed the remaining 372 videotapes, whch contained 4976 decisions. P8L41-3: provide either SD or CI, both is redundant.
P9 L5-12 are confusing to a naïve reader. Please start with "Table 3 shows that categories 1-19 etc, and end at "being most frequent."
P9, L14, and further in the doc: Do not start a sentence with a number, but write in full. Here: Eighty-one P9, L 18-21: all these procedures are not relevant to the reader, so I suggest writing "(21%), and other (all less than 10%)" or something similar.
P9, L45 Table 5 refers to a research question that was not posed, in my view. Something like assess association with specialty. Further, the temporal categories are part of the classification, wo
P10, L3-6 suggest to add: "and the "here and now"decisions was significantly lower than"
Table 6: when judging the columns (or deciding about them, as you call it $\Box$ ) it would have been nice for me as reader to have the 16, 73 and 13% in the header for the 3 temporal types, but perhaps this comment becomes irrelevant after my suggested changes above.
P10, L20: low numbers of advice and precautions in Wards. Were there no patients who would be dismissed that day then? For that's when one expects these decisions (or recommendations as I would call them $\Box$ ) to be communicated to the patient a lot. Perhaps a comment in this respect in the Discussion?
P10,L 34, this paragraph is a bit tedious, and I am not sure it adds much for the reader? I suggest deleting it or making it a list in an Appendix. Besides, what do you mean by positive evaluations of test results, and interpretation as satisfactory? Was the test performed well? Or was the result in the healthy range?
P11, L 12: gender, age or setting, is that of patients?
P13 L 14-18. I do not see why this supports the validity, had the results been different, another argument would have held, and the validity also, I fear. I suggest not concluding that this is a validation.
<ul> <li>P13, L 49+: Given my comments, I suggest altering this paragraph. I do not see how the findings in the tables can help clinical studies and normative and prescriptive judgments of practice. Please elaborate on what is meant here. (by the way, I am not sure that in English the plural of diagnosis and prognosis would be used here?). The discussion of the normativity in the manuscript to me simply is not clear.</li> <li>What is a framework exceeding the encounter?</li> <li>P14, L20-22: what is meant by this first sentence ("Introducing physicians communicated")? As an intervention tool to improve communication? Please explain.Similarly in the next paragraph, about putting the punctuation marks out in the open.</li> </ul>
P15 Conclusion: What do you mean by your last sentence? That people should do this with your system? It is not what you have done here, provided an exhaustive description of how decisions are communicated. So why is this a conclusion?

REVIEWER	Karen Sepucha, PhD Massachusetts General Hospital, Harvard Medical School USA Dr. Sepucha receives salary support as a member of the scientific advisory board for Healthwise, a not for profit company that produces patient educational materials including patient decision aids.
REVIEW RETURNED	08-Aug-2017
GENERAL COMMENTS	In this manuscript, the investigators cataloged the decisions that were made during the course of 380 videotaped hospital encounters. The encounters were with 372 patients and 58 physicians covering a range of specialties. The authors used DICTUM a framework that they developed with this data set and published previously. The quantity of decisions found in these short encounters was significant—an average of 13. The well-written descriptive study provides insight into the type of activities within these encounters, and the range of judgments and decisions being made by clinicians. The enthusiasm for the paper is dampened by the taxonomy the authors used that conflates observations, explanatory statements, medical opinions/judgments and decisions, and that discounts the patient or family in the process.
	Major concerns: 1. More than half of the "decisions" were in categories that were solely statements of explanation or conclusions, without requiring consideration of options or action (both of which are pre-requisites in definitions of decisions). One major concern is the very broad definition of decision Pg. 5 "any statement concerning the patients' health that carries meaning and weight because it is said by a medical expert counts virtually every statement made by the doctor as a decision. By including observations (e.g. your blood pressure was 180/100) and conclusions based on these observations (e.g. your blood pressure is high), the authors are diluting the definition of decision and confusing it with observations and explanatory statements. 2. Although they are not decisions, these types of observations and judgments are important to highlight when studying decision making. However, it would be better to categorize the explanatory statements (such as explaining test results, describing diagnoses, and clarifying treatment targets or goals), as something other than decisions. In fact, the authors even describe these differently (e.g. 2 and 3 in Table 1) in their taxonomy. It would be important for the authors to be clearer conceptually on their definitions and theoretical foundation for their definition. Further, please clarify how explanatory statements (your blood pressure is 180/100) are decisions. 3. Another major problem with their approach is that physicians who take time to explain their reasoning and rationale would be counted as having made more decisions, when in fact they might just be communicating more comprehensively with patients. A statement by a doctor, "You are looking well today. Your blood pressure is under control. I think it is time to discharge you." Which would (I think) be counted as potentially three decisions in their taxonomy. 4. It is also problematic that patients' judgments and observations (for example, which symptoms to share, or their opinion

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

This is a major gap that is at adds with more gaparal approaches to
This is a major gap that is at outs with more general approaches to
patient engagement and shared decision making that values the
patients' expertise and contribution to decision making.
5. The authors categorized the timing of the decisions and it appears
that they counted 'pre-formed' decisions as happening in the current
encounter. Were these decisions revisited and meaningfully
discussed or were they merely stated? If just stated, then it does not
seem like these should count as decisions being made in the current
ancounter. That would make a significant difference.
200/ of the word decisions
39% of the ward decisions.
Minor:
-Why was the physician with long consults excluded? It is possible to
code those visits—perhaps those were significant issues (e.g. new
cancer diagnosis) that would require additional time. Is this
framework only relevant for short consults?
-It would be helpful if the authors would add some details as to
whether coders were watching the video or coding a transcript of the
ancounter
encounter.

# **VERSION 1 – AUTHOR RESPONSE**

Reviewer: 1 Reviewer Name: Anne Stiggelbout Institution and Country: Leiden University Medical Center, Netherlands Competing Interests: None declared

Comment: I am not a native speaker, and neverteless found some phrases (e.g. amount of decisions instead of number, less instead of fewer, a statement that is said) that do not sound right, so I suggest having a native speaker check the english. I miss some comma's and articles as well, that would increase readability.

Response to Reviewer: As mentioned, the manuscript has been thoroughly proofread by a native English-speaking researcher working within the field of communication in health care. Changes are found in blue colored text in the revised submission.

Comment:And sentences should not start with a number (a percentage, e.g. P9, L 14, P10, L 52 and 54)

Response to Reviewer: Thank you for this comment. All sentences that started with a number in the original submission, have been adjusted.

See further my comments in file, and comments in the pdf of the manuscript:

Reviewer 1 comment file:

I commend the authors for their extensive work on classifying decisions, particularly with these large numbers. Most of my queries/questions, relate to their definition of a decision and their classification, which is somewhat difficult because it has already been published by BMJOpen. So most of my comments are textual, of necessity.

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Even though the classification is out in the open, I have been googling, and depending on the website accessed, decisions and judgments are treated as almost the same or different.

The authors build on Braddock: "a decision is a verbal statement committing to a particular course of action", and have simply added: "and/or statement concerning the patient's health that carries

meaning and weight because it is said by a medical expert". This to me opens up the definition to any statement by a doctor concerning the patient's health (even false statements!). So then a judgment, a recommendation, anything related to the patient's health becomes a decision? Doctor:"mr Jones, you are obese" now is a decision? With which I personally disagree.

The example given on blood pressure exemplifies this: "Your blood pressure is 180/100. This is high." So the first statement is a fact, the second a decision (according to the authors). I would consider this a judgment. "Given your age, this means: high." For me the decision is: "we need to do something about this" (or even better, "we have three options, do nothing, change behaviour, take pills", and which one is decided upon to me would be the decision).

Some people for this reason define a Decision as "a choice between options". Of course, one may say that High and Not High are two options, but then anything we like in life is a decision (e.g. "This is an excellent book!").

Personally I like the distinction from the Lonergan Forum:

Judgment: "complete one's knowledge of an actuality that already exists" vs. Decision: "confer actuality upon a course of action that otherwise would not exist". So the judgment is that the patient has high blood pressure (compare his level with a standard), the decision is that he should take pills.

Response to Reviewer: Reviewer 1's thoughts about this precisely describes the core of the challenges we had when aiming to identify and classify all clinically relevant decisions. Braddock et al. defined decisions as "verbal statements committing to a particular course of action", which is a good starting point. What we have added to this definition, is that the clinically relevant judgments that medical actions are built upon also are a result of decision-making processes and that they are equally relevant. If the judgment is wrong, the actions recommended as a result might be equally wrong. And yes, even false statements would be registered – and possible to assess with a normative approach, in studies where normative assessment of clinically relevant decisions would be the objective.

To build on Reviewer 1's example on whether "You are obese", the context in which this statement is made, our criteria presented in the Methods section and the codebook would provide guidance as to whether we would code is as a clinically relevant decision. If these words are stated at an obesity clinic that the patient has been referred to in order to get medical care and it is obvious to all parties involved; the patient and the physician – we would not include it. The reason for not including it is found in our definition: "a verbal statement committing to a particular course of clinically relevant action and/or statement concerning the patient's health that carries meaning and weight because it is said by a medical expert." Stating the obvious does not provide new meaning.

However, if the patient goes to see an orthopedic surgeon because of pain in a knee and the orthopedic surgeon was to say: "Your main medical problem is that you are severely overweight", implying that weight loss would alleviate the pain, we would code that as a clinically relevant decision. Our point is that clinical judgments are confronted with options (high/low) and preferences (of both patients, physicians and medical care systems). If medical problem-solving tasks always were preference-independent, we would not create any discussion about this. But when the combination of context and uncertainty is involved, preference comes creeping into the equation. Our example of 180/100 is not about what the which numbers the test yields, it is about what those numbers mean, in this context. Blood pressure-measurements do not speak for themselves. Somebody has to interpret what a blood pressure of 180/100 means in a given context. Most of the time this is an easy task, because 180/100 is a high blood pressure and in most contexts it could be judged as that.

But in other contexts, for instance in the context of a massive stroke or in a patient with a hypertensive crisis, 180/100 could be a blood pressure that physicians – for the time-being might be happy with, it could even be regarded as ideal for the given context and time frame.

So we argue that what has previously been framed as mere problem-solving tasks can in fact involve preferences, including what outcomes matter to the patient (and physician), and how they feel about risks and benefits (of this blood pressure, in this context). What roles patients wish to play in decision-making with their physicians – and what roles physicians want to play and let patients have - is a key interest within our research group. And we agreed that in order to pursue this interest, we first needed a broad, precise and comprehensive description of how decisions are communicated in these encounters by identifying and classifying the punctuating conclusions of decision-making dialog, whether or not they are delivered as paternalistic information or reached by collaborative deliberation. The development of DICTUM has been a stepping stone to our current work, in which we assess how subsets of decision-making discussions align with what the literature frames as state-of-the art communication involving patients in medical decisions.

We believe that a detailed and precise description of decision statements – both qualitatively and quantitatively – is relevant both to physicians and patients. From a medical point of view, the taxonomy comprises any clinically relevant task that needs to be dealt with in an encounter; from interpreting the patient's story, symptoms, clinical findings and diagnostic tests, to the translation of this knowledge into actions including medical interventions, providing relevant contextualized information to the patient and appropriate level of follow-up.

From the patient's perspective, the statements coded as decisions sum up bullet points of information the patient can take home from the encounter. Imagine a patient coming home to his spouse or parent and being asked; "So what did the doctor say?" The response could be a summary of the statements identified as decisions by the taxonomy, e.g. "The doctor concluded that I have pneumonia and gave me some antibiotics. She said I will be fine again, but that it could take as long as a month before all symptoms will pass. I have to go back to control my chest x-ray in 6-8 weeks. She said I should stop smoking. When I asked if I could get any of the pills available for smoking cessation, she said I have to speak with my family physician." This example is probably more structured, detailed and medico-centered than patients' real-life summaries of medical encounters would be, but it is provided to depict the amount and complexity of clinically relevant outcomes (judgments and actions) that is communicated to patients.

Comments: The authors have published this classification, so I would like to handle the manuscript as such in my review.

1. My major comment then is that a comment in this respect on their own classification is needed in the Discussion, for the high number of decisions are to a large extent due to this inflation of the term Decision. The authors provide this as a limitation (P13, L 20-47), but the argument that other studies find lower numbers because they only looked at patient involvement is not the appropriate one, I would argue. It is simply the inclusion of judgments or statements in the definition. Diagnostic decisions should in my view be decisions about diagnostics, in which one may or may not involve a patient ("I think you should have an MRI, do you agree?") and not the provision of test results per se.

Response to Reviewer: We agree that the main reason why the number of decisions is higher in our study is because we have included clinically relevant judgment statements in the taxonomy and subsequent analysis of our material. We have deleted previous text about measurement of patient involvement being a reason for lower numbers in previous studies: "and the selection of decisions appears limited by this aim" P4 L56-P5 L3

We agree that the provision of test results per se does not qualify as a clinically relevant decision, but that the interpretation of test results in the actual patient's context is a key conclusion that will guide any (discussion about) subsequent medical actions.

Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies

2. On page 5, they describe that in theory the distinction problem solving and decision making is sensible, but that in clinical practice the two are constantly blurred. Since this is a major argument for their broad definition of Decision, I would like to have more substantiation for this statement.

Response to Reviewer: We agree. We have deleted the final part of the sentence "...constantly blurred" and elaborated on this in two sentences and with two new references (reference 22 and 23).

3. And I do not understand the subsequent statement that the normative and prescriptive approaches "need a descriptive framework". Please explain to a larger extent in the text.

Response to Reviewer: We see how this can be unclear and have added a sentence with hopes to increase clarity.

4. Please motivate the addition of "and/or statement concerning the patient's health that carries meaning and weight because it is said by a medical expert". Readers should not need to go back to the 2016 paper in BMJ Open.

Response to Reviewer: By answering to Reviewer 1's point number 2 and 3 we feel that our motivation – finding out what judgments clinical actions/recommendations/proposed tasks are based upon – has been more explicitly disclosed.

5. I think Table 3 does not add much, can be in an Appendix.

Response to Reviewer: We agree that Table 3 does not add a lot: it is neither precise, nor detailed, but it lends an indication towards the diversity of the material. We will happily move it to be an Appendix instead.

6. The topical categories should also be in Table 4, as they are in Table 1 and in the classification treated similarly to the topical ones. I feel it is therefore also not logical to treat the temporal ones differently from the topical ones in Tables 5 and 6. Logical would be if T5 had the columns Ward, Outpatient, ER, and then first the topical ones and below that the temporal ones, as in Table 1. It is odd to have them as columns in T6, as they are part of the classification and not independent factors. See what I mean?

Response to Reviewer: Yes. We see what you mean. Thank you for the advice. As a result, we have added the numbers describing the temporal categories from table 5 to table 4 and table 6, and deleted table 5. As we have moved table 3 to the Appendix the table presenting topical and temporal categories (total number, present in number of encounters, average and min-max) is the new table 3 and the table presenting topical and temporal categories in three different clinical settings (outpatient, ward round and emergency room) is now table 4.

7. Table 7 can be deleted, does not provide much insight, and the only significant numbers presented in the text. It feels a bit like a fishing expedition. What's the rationale or theory behind the chosen independent variables? Further, given the very broad definition of decision, the interpretation is cumbersome of the differences between specialties. And are these the results from Bivariate or Multiple models? Temporal by specialty might be more interesting? Or the results of this table could perhaps be presented in a figure like Fig 1 (and instead of that one, which is confounded by e.g. specialty). The section in the Discussion, on P12, L 27-52, makes it seem as if there was some hypothesis beforehand, rather than fishing, but on the other hand, these arguments could have been easily altered had something different been found. E.g. why do ENT encounters commonly deal with only one concern?!

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

Response to Reviewer: We would like to keep table 7 (now named table 5 as of the adjustments mentioned above). We understand that us going deep into the statistics with table 7 could seem motivated by fishing, but truly it was not. When we were fortunate to have a large material from many specialties, analyzing the average number of decisions per specialty was done to describe the landscape of hospital work. We did not have any hypotheses of differences between specialties. The striking result here is the lack of difference between age groups, gender and specialties. Where we found minor differences, we tried to suggest medically informed common sense explanations to what we found. Which to us is what the Discussion section should be about: trying to interpret your findings and critically evaluate possible explanations for them.

Why do ENT encounters commonly deal with only one problem? Because they do. ENT encounters can of course can deal with several problems, for instance in ENT related oncology or complex cases with cochlea implantations and neuro surgery, but in our material of everyday ENT encounters, they had an average of seven decisions (according to our definition, taxonomy and coding), which we interpreted to be caused by a limited number of clinical problems addressed. The reason for estimating linear mixed models accounting for random and fixed effects, was upon request from a reviewer for Medical Decision Making. Originally, we only compared means between two groups using independent sample t-tests and between several groups using one-way analysis of variance (ANOVA) with Bonferroni correction.

8. Since the readers should not need to go back to the 2016 paper, I would like to see a comment on the coding of the Preformed decisions. A limitation is that the coder cannot always no if the decision was preformed. E.g. in oncology, many decisions have been preformed in the Multidisciplinary Team Meetings, but this may not be stated as such to the patients. The doctor will simply say: "we". (See Engelhardt et al. EJC 2016, page 59, left column, second para). So I am convinced there will have been more preformed than counted. Please comment on this in your Discussion. This is stated on P12, L 34, for WR decisions, but holds for many specialties nowadays too.

Response to Reviewer: We agree that preformed decisions is hard to be certain about from an observator's perspective. We have added a sentence about this in the limitations paragraph of the discussion. And we agree that in oncology, preformed decisions are particularly frequent as modern oncology heavily relies on the input from MDT meetings (which in turn might make shared decision-making challenging, but that is a another, extremely interesting phenomenon to study, that the first author have concrete plans to study in the future).

## Textual:

P3 Not all readers will see why the statistical analysis is a strength of this study, please explain. Response to Reviewer: We agree. Seeing that this material comprises 380 encounters spread across 58 physicians, we needed a statistical model to ensure that observed differences were not attributable to significant clustering at doctor level. We have added a few words to the sentence to improve precision and clarity.

P4 L 11-12 Since Finestein's conclusion was drawn in 1994, I would change "has left" in line 11-12 to "had left"

Response to Reviewer: We agree. We have changed it to "had left".

P4 L 16-21: the paper mostly speaks of decisions to do something, I suggest adding –e.g. in this paragraph- doing nothing also as an outcome of a decision. Also e.g. p5, L11-16.

Response to Reviewer: We agree that deciding to do nothing is also an important outcome and have added it to the sentence on P4, L16-21.

P4 L 21 I do not see why this is "but", the pressure is not necessarily related to who makes the decision. I suggest rephrasing this sentence.

Response to Reviewer: We have removed the word "but". We agree that who makes the decision is not the only reason for the change we have seen in this field the past decades.

P4 L34-39, EBM for a long time now has also included patient preferences in addition to best available evidence!

Response to Reviewer: We are aware of patient values and preferences being added to the definition of EBM the latter years. Because it was not a part of the original descriptions of EBM and because SDM promotes patient preferences and values in decision-making much more explicit, we did not include it in the original submission. But we have inserted it at the end of the sentence about EBM at

### P4 L34-39.

P3 L45: Research and implementation not only target single decisions, suggest adding Often. Further, I disagree with the phrasing "addresses decisions where medical evidence provides no clear guidance", to a naïve reader this may sound as if SDM is only for cases where the doc does not know. But the clear guidance may be that there are two options that patients may weigh differently and thus the patient should be involved...

Response to Reviewer: We have added the word "often" to the sentence and changed the end of the sentence from "addresses decisions where medical evidence provides no clear guidance" to "... two or more options that patients may weigh differently".

P6, L32: please explain Broad consent at its first use.

Response to Reviewer: We have added two sentences about how the consent for the 380 encounters included in this study was broader than it was for the 497 encounters included for analysis in the original RCT.

P6, L 40-45: the response should be in the Results section. And please provide the 58% response after the 59 who provided broad consent.

Response to Reviewer: Because the "response rate" of physicians and patients "belongs" to a previous trial, we originally described it in the Methods section, but we have moved it to the first paragraph of the Result section on Reviewer 1's request.

P7 L 32: please explain how intra-rater reliability was assessed. Did EHO code x video's twice, and how long was the period between the two?

Response to Reviewer: EHO coded the videos for intra-rater reliability approximately one year after the initial coding of the same videos. We have inserted this information in the Methods section.

7 L 43=48: are "participants' patients? And please check parentheses, after old patients, and physicians. I suggest disentangling patients and physicians, for clarity.

Response to Reviewer: Participants means both patients and physicians. We have specified this in the revised version of the manuscript. We have removed the text within the parentheses as Table 1 more precisely shows how patients and physicians were stratified in relevant age groups.

P7 L52: apparently the aim of the study was also to compare decisions within categories. Please add this to the aim of the study, in abstract and on P5, otherwise this sentence and the analyses on page 8, do not make sense.

Response to Reviewer: We have added a sentence to the abstract and in the final paragraph of the Introduction on page P5.

P8, L 38-39: We analysed the remaining 372 videotapes, which contained 4976 decisions. P8L41-3: provide either SD or CI, both is redundant.

Response to Reviewer: We agree that reporting both is redundant. We have deleted the CI.

P9 L5-12 are confusing to a naïve reader. Please start with "Table 3 shows that categories 1-19 etc.., and end at "being most frequent."

Response to Reviewer: We agree. As we due to Reviewer 1's request have moved Table 3 to the appendix, we start the sentence with "The Appendix table..."

P9, L14, and further in the doc: Do not start a sentence with a number, but write in full. Here: Eightyone...

P9, L 18-21: all these procedures are not relevant to the reader, so I suggest writing "(21%), and other (all less than 10%)" or something similar.

Response to Reviewer: We have rearranged the sentence starting with a number. Thank you. And we have deleted all procedures accounting for less than 10% of the procedures observed.

P9, L45 Table 5 refers to a research question that was not posed, in my view. Something like assess association with specialty. Further, the temporal categories are part of the classification, wo

Response to Reviewer: Table 5 has been deleted as a part of Reviewer 1's suggestions for improvement earlier in the review. We hope this meets with Reviewer 1's wishes for this commentary point.

P10, L3-6 suggest to add: "and the "here and now" decisions was significantly lower than..."

Response to Reviewer: Thank you. We have added "decisions was significantly lower than..."

Table 6: when judging the columns (or deciding about them, as you call it<sup>(2)</sup>) it would have been nice for me as reader to have the 16, 73 and 13% in the header for the 3 temporal types, but perhaps this comment becomes irrelevant after my suggested changes above.

Response to Reviewer: We have altered the tables according to Reviewer 1's suggestions above and hope this meets with expectations for readability and clarity.

P10, L20: low numbers of advice and precautions in Wards. Were there no patients who would be dismissed that day then? For that's when one expects these decisions (or recommendations as I would call them<sup>(C)</sup>) to be communicated to the patient a lot. Perhaps a comment in this respect in the Discussion?

Response to Reviewer: We agree that this number was low and we agree with Reviewer 1's expectations for a higher number in discharge encounters. Of the 58 ward round encounters less than ten were discharge encounters.

P10,L 34, this paragraph is a bit tedious, and I am not sure it adds much for the reader? I suggest deleting it or making it a list in an Appendix. Besides, what do you mean by positive evaluations of test results, and interpretation as satisfactory? Was the test performed well? Or was the result in the healthy range?

Response to Reviewer: As a part of a landscape description, we included this paragraph to provide further detail and precision into subcategories of the topical categories. EHO originally presented this information a large table.

Since Reviewer 1 finds the paragraph tedious and adding little for the reader, we have deleted it. Satisfactory test result interpretations means that the test was (decided by the physician) to be within a healthy range.

P11, L 12: gender, age or setting, is that of patients?

Response to Reviewer: Of patients and physicians, we have clarified this. Thank you.

P13 L 14-18. I do not see why this supports the validity, had the results been different, another argument would have held, and the validity also, I fear. I suggest not concluding that this is a validation.

Response to Reviewer: We have presented our findings to clinicians in more than 20 different for the past three years and clinicians nod and says that our findings support their experience of clinical work and decision-making. But since Reviewer 1 disagrees with the validity of our findings, we have deleted the sentence.

P13, L 49+: Given my comments, I suggest altering this paragraph. I do not see how the findings in the tables can help clinical studies and normative and prescriptive judgments of practice. Please elaborate on what is meant here. (by the way, I am not sure that in English the plural of diagnosis and prognosis would be used here?). The discussion of the normativity in the manuscript to me simply is not clear.

Response to Reviewer: We agree that the findings in the tables will not be of much use to clinical studies and normative and prescriptive judgments of practice. But behind every number is a statement vocalizing a judgment or a recommended action. This study shows that all these statements can be identified, classified and that they spread out like we have shown in a large number of hospital encounters. The first author of this paper is currently doing a study to assess the quality of decisions in medical encounters and is using DICTUM to identify clinically relevant decisions, judgments and actions alike, before interviewing patients and physician about their perspectives on the decision-making processes. Knowing what to look for and how to describe it is a key element in order to assess the quality of it. What is a framework exceeding the encounter?

P14, L20-22: what is meant by this first sentence ("Introducing physicians.. communicated")? As an intervention tool to improve communication? Please explain. Similarly in the next paragraph, about putting the punctuation marks out in the open.

Response to Reviewer: We see how this could be unclear. We have added words to make it explicit that a framework assessing decisions may exceed the duration of the isolated encounter (e.g. series of encounters during a hospital stay, follow the physicians to meeting where their patients are discussed). In P14 L20-22 we mean that it would be interesting to see if introducing patients and physicians to the taxonomy and its topical and temporal categories, would change how they discuss decisions or not. We would not use the term intervention tool, because we would know what to expect from it. But for starters, as a small experimental study, it would be interesting. Clinical decisions, both judgments and actions, can be described as punctuation marks in medical encounters. Raising awareness around decisions, would make decisions more explicit to both patients and physicians – and hopefully together with other resources (e.g. decision aids, communication tools for SDM) increase patient involvement in decision-making processes concerning their own health.

P15 Conclusion: What do you mean by your last sentence? That people should do this with your system? It is not what you have done here, provided an exhaustive description of how decisions are communicated. So why is this a conclusion?

Response to Reviewer: We set out to develop a descriptive tool to more precisely know what we would be assessing, if we were to continue with assessing the quality of medical decisions. We feel that we have developed a descriptive framework that identifies and classifies clinically relevant decisions with precision and detail, and have in the process understood that to assess quality of the decisions we also need input from the patient and physician perspective.

#### Reviewer: 2

Reviewer Name: Karen Sepucha, PhD

Institution and Country: Massachusetts General Hospital, Harvard Medical School USA Competing Interests: Dr. Sepucha receives salary support as a member of the scientific advisory board for Healthwise, a not for profit company that produces patient educational materials including patient decision aids.

Comment: In this manuscript, the investigators cataloged the decisions that were made during the course of 380 videotaped hospital encounters. The encounters were with 372 patients and 58 physicians covering a range of specialties. The authors used DICTUM a framework that they developed with this data set and published previously. The quantity of decisions found in these short encounters was significant—an average of 13. The well-written descriptive study provides insight into the type of activities within these encounters, and the range of judgments and decisions being made by clinicians. The enthusiasm for the paper is dampened by the taxonomy the authors used that conflates observations, explanatory statements, medical opinions/judgments and decisions, and that discounts the patient or family in the process.

Response to Reviewer: We thank Reviewer 2 for the appreciation of the insights into hospital encounters provided by our study. We are sorry to see that we have not been clear enough about what category 2 and 3 codes in our taxonomy (evaluating test results and defining problem) consist of. We have tried to increase clarity on this throughout the paper and will respond more specifically to Reviewer 2's concerns below.

Major concerns:

1. More than half of the "decisions" were in categories that were solely statements of explanation or conclusions, without requiring consideration of options or action (both of which are pre-requisites in definitions of decisions). One major concern is the very broad definition of decision Pg. 5 "any statement concerning the patients' health that carries meaning and weight because it is said by a medical expert counts virtually every statement made by the doctor as a decision. By including observations (e.g. your blood pressure was 180/100) and conclusions based on these observations (e.g. your blood pressure is high), the authors are diluting the definition of decision and confusing it with observations and explanatory statements.

Response to Reviewer: As we have replied to Reviewer 1, we agree with Reviewer 2 that by broadening the identification and classification to include also clinically relevant judgments, we include more statements than if we had only identified action statements. We do not necessarily agree that we are diluting the term of clinically relevant decisions, rather we argue that the taxonomy adds detail and precision to decisions leading to actions, by also describing the rationale behind them.

"Your blood pressure is 180/100" is not a decision. The decision is whether or not that blood pressure is judged as good/normal/acceptable or bad/in need of treatment. If this judgment is wrong, the action to treat/not treat the blood pressure will be equally wrong. We have been more explicit about the taxonomy including both clinical judgments and clinical actions throughout the paper.

2. Although they are not decisions, these types of observations and judgments are important to highlight when studying decision making. However, it would be better to categorize the explanatory statements (such as explaining test results, describing diagnoses, and clarifying treatment targets or goals), as something other than decisions. In fact, the authors even describe these differently (e.g. 2 and 3 in Table 1) in their taxonomy. It would be important for the authors to be clearer conceptually on their definitions and theoretical foundation for their definition. Further, please clarify how explanatory statements (your blood pressure is 180/100) are decisions.

Response to Reviewer: We are glad that Reviewer 2 agrees with us that clinical judgments are important to highlight when studying decision-making. As we have responded to Reviewer 1, interpretations (not explanations) of test results, deciding on (not describing) diagnoses and deciding on where to set treatment goals require the same kind of cognitive effort as deciding on relevant clinical actions. Diagnoses most commonly do not reveal themselves. Of course, anyone can see that if a bone pipe is sticking out of the shin, the leg is broken. But most of the times deciding on a diagnosis requires making choices between options (what information to weight and what information to discard) and to conclude taking various levels of uncertainty and ambiguity into account.

3. Another major problem with their approach is that physicians who take time to explain their reasoning and rationale would be counted as having made more decisions, when in fact they might just be communicating more comprehensively with patients. A statement by a doctor, "You are looking well today. Your blood pressure is under control. I think it is time to discharge you." Which would typically be considered one decision with some rationale, would (I think) be counted as potentially three decisions in their taxonomy.

Response to Reviewer: Reviewer 2 is right that for the example presented we would not only look for the action (discharge), but also for the judgment behind why this action is considered the appropriate one. As we replied to Reviewer 1, from the patient's perspective, the statements coded as decisions are bullet points of information the patient can take home from the encounter. If the patient is to call home and tell about the pending discharge, his/her spouse will most likely get, and probably would want, the rationale behind that decision (e.g. the blood pressure is now under control).

In our current study of patient involvement in medical decisions, the identification of both the proposed clinical actions and the clinical rationale behind them is essential to understand how physicians and patients communicate and make decisions, and to a large extent explain why patients do not get involved as much as we (researchers of communication in health care) would want. More on that in a soon to be submitted paper.

4. It is also problematic that patients' judgments and observations (for example, which symptoms to share, or their opinion of what might be causing a particular problem, or their judgment about which approach might be best) are not counted within this framework. This is a major gap that is at odds with more general approaches to patient engagement and shared decision making that values the patients' expertise and contribution to decision making.

Response to Reviewer: We agree with Reviewer 2 that the aim of identifying clinically relevant decisions communicated in patient-physician dialog, has led to a taxonomy that is more physician-centered, than patient-centered. We have tried to be explicit about this all the way from the title (Clinical decisions presented to patients...). But it is important to add that patients' opinions about diagnosis, prognosis and etiology or interpretations of test results - if stated in the encounter – also results to the identification and coding of clinically relevant decisions.

5. The authors categorized the timing of the decisions and it appears that they counted 'preformed' decisions as happening in the current encounter. Were these decisions revisited and meaningfully discussed or were they merely stated? If just stated, then it does not seem like these should count as decisions being made in the current encounter. That would make a significant difference—e.g. removing 39% of the ward decisions.

Response to Reviewer: Preformed decisions account for a lot of the statements coded as decisions, especially on ward rounds. They have been made before the encounter and are presented as information to patients and not discussed further unless the patient initiates talk about them. In our view, this is an important – and partly disturbing – finding, and in a previous paper we have described how this way of making decisions in hospital has the possibility to exclude the patient from the point of decision-making. (Ofstad EH, Frich JC, Schei E, Frankel RM, Gulbrandsen P. Temporal characteristics of decisions in hospital encounters: A threshold for shared decision making? A qualitative study. Patient Educ Couns. 2014; 97(2):216-22.)

We could have discussed our findings with a patient-centered or shared decision-making perspective, but have tried to focus on being descriptive, and not normative or prescriptive. The taxonomy is our stepping stone to go further and approach clinical decision-making with a normative and prescriptive perspective – which is and will continue to be the main focus of our research group.

## Minor:

-Why was the physician with long consults excluded? It is possible to code those visits—perhaps those were significant issues (e.g. new cancer diagnosis) that would require additional time. Is this framework only relevant for short consults?

Response to Reviewer: Why the outlier was an outlier is no mystery, but because more detailed information could have exposed her identity, we did not elaborate on this. The physician we excluded was unexperienced and in the final trimester of pregnancy. Her seven encounters (all emergency department encounters) were on average 4,5 times as long as the average of other encounters, the longest lasted 3 hours.

The seven encounters accounted for 23% of all emergency room encounters and dealt with the same clinical problems as the remaining 23 encounters (i.e. question about deep vein thrombosis, chest pain and shortness of breath). The physician today works in a tempo that is comparable with the other physicians in the study.

The longest encounter included in the analysis was 66 minutes and the framework is relevant for encounters of all durations.

-It would be helpful if the authors would add some details as to whether coders were watching the video or coding a transcript of the encounter.

Response to Reviewer: Sorry if this was unclear. Analysis was done through video observation, not transcripts. We have added a sentence about this in the Methods section.

## **VERSION 2 – REVIEW**

REVIEWER	Anne Stiggelbout Leiden University Medical Center
REVIEW RETURNED	04-Oct-2017

GENERAL COMMENTS	I am glad the authors decided to add the term judgment in many
	uploaded word file, for generally judgments precede actions and decisions
	The reviewer also provided a marked copy with additional comments. Please contact the publisher for full details.

REVIEWER	Karen Sepucha Massachusetts General Hospital, Harvard Medical School, USA Dr. Sepucha receives salary support as a member of the scientific
	advisory board for Healthwise, and not for profit that creates patient educational materials.
REVIEW RETURNED	10-Oct-2017
GENERAL COMMENTS	The authors did make several changes to address my comments and those of the other reviewer. However, as is clear with both reviewers' comments, a main challenge with this paper is their taxonomy and in particular, their expanded definition of "decisions." The taxonomy has introduced many conceptual challenges and confusion (e.g. counting explanatory statements and observations as decisions, though only when made by doctors and not patients, counting pre-formed decisions as ones made in the current encounter, etc.). This led to an inflation in the number of decisions that are counted within each encounter, and skewed the results. Also, in their response to comment 4, they argue that only doctors can identify clinical relevant decisions but then also state that, "But it is important to add that patients' opinions about diagnosis, prognosis and etiology or interpretations of test results - if stated in the encounter – also results to the identification and coding of clinically relevant decisions." The response is confusing, did patients' comments count as clinical decisions? That is not accurately described in their methods or coding.

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies

In reading the revision, I still have the same major concerns as with the initial paper. Given that they are not presenting this taxonomy here, it is not clear how to handle this fundamental disagreement. It
may rest on the editor to decide whether the paper, using the
taxonomy as published previously, has a net benefit for readers.

### **VERSION 2 – AUTHOR RESPONSE**

Reviewer: 1 Reviewer Name: Anne Stiggelbout Institution and Country: Leiden University Medical Center Competing Interests: None declared

Comment: I am glad the authors decided to add the term judgment in many instances. I have suggested a change in order, though in the uploaded word file, for generally judgments precede actions and decisions....

Response to Reviewer 1: Thank you for your positive evaluation of our work. We have inserted the wording "judgments and actions" at the three spots indicated by Reviewer 1, and also in the abstract. We have added the p-value value for drug-related decisions as requested. And we have divided the sentence with inter- and intra-rater reliability into two separate sentences for clarity.

### Reviewer: 2

Reviewer Name: Karen Sepucha

Institution and Country: Massachusetts General Hospital, Harvard Medical School, USA Competing Interests: Dr. Sepucha receives salary support as a member of the scientific advisory board for Healthwise, and not for profit that creates patient educational materials.

Comment: The authors did make several changes to address my comments and those of the other reviewer. However, as is clear with both reviewers' comments, a main challenge with this paper is their taxonomy and in particular, their expanded definition of "decisions."

The taxonomy has introduced many conceptual challenges and confusion (e.g. counting explanatory statements and observations as decisions, though only when made by doctors and not patients, counting pre-formed decisions as ones made in the current encounter, etc.). This led to an inflation in the number of decisions that are counted within each encounter, and skewed the results.

Response to Reviewer 2: We note that Reviewer 2 thinks our taxonomy introduces confusion. Our intention has always been the opposite, to further explore and increase our understanding of decision-making in clinical encounters. We started this project seven years ago with the intension of applying previously developed definitions and tools to identify and classify decisions. As much as we look up to and have been inspired by the pivotal contribution to this field, namely Clarence Braddock's work 15-20 years ago, we - as physicians with clinical experience – soon realized that defining decisions merely as actions, leaves out the foundation and the very essence and decisional processes these actions are based upon. And as our ultimate goal has been to assess clinical decisions (the quality of the decisions, from an evidence-based, patient-centered, cost/benefit and public health perspective), we could not approach this question without trying to understand more about the various decisions that is being made in clinical encounters.

By defining decisions both as judgments and actions, the number of decisions identified have increased compared to previous studies. An increase is, in our eyes, not an inflation, but an enhancement of detail and precision. We disagree that our approach skews the results. We are explicit about our tool and the reason for the increase in number of decision. When our findings are broken down into categories, they actually validate previous studies that have applied a narrower definition of decisions. We highlight this in the Discussion: "One may challenge our definition of decisions. Previous studies of decisions in patient-physician encounters have reported substantially lower frequencies, varying between, on average, three and seven decisions per encounter in five different studies. (15-19) Each of these studies have identified decisions with the aim of describing patient involvement in decisions. These studies did not include diagnostic decisions (comprised by our first three categories); if diagnostic decisions are subtracted from our material, our findings align with the findings from previous studies."

As pointed out in the last response letter to the reviewers, we do not code observations as decisions. What we code is the interpretation of observations, uttered by the physician either as simple evaluations (e.g. "Your blood sugars and your A1c are right where we want them") or as more complex interpretations of several simple evaluations (e.g. "I think you have got a pneumonia" derived from the interpretation of patient story, clinical examination (fever or not) including auscultation (crackles or not), chest x-ray (infiltration or not), laboratory tests (increased white blood count or not, increased C-reactive protein or not) etc).

And we do not code explanatory statements as decisions, unless they concern the patient's diagnosis ("the name of the beast"), it's etiology ("where it comes from") or it's prognosis ("how long it will be here and how it will affect you"). Behind these statements, there is always a cognitive process in which options and preferences present themselves, and the person making these statements are confronted with choices every way of this process.

Reviewer 2: Also, in their response to comment 4, they argue that only doctors can identify clinical relevant decisions but then also state that, "But it is important to add that patients' opinions about diagnosis, prognosis and etiology or interpretations of test results - if stated in the encounter – also results to the identification and coding of clinically relevant decisions." The response is confusing, did patients' comments count as clinical decisions? That is not accurately described in their methods or coding.

Response to Reviewer 2: Thank you for this comment. Patients make decisions, both as judgments and actions. Sometimes these judgments (e.g. "I think I have pneumonia") and actions (e.g. "I want to start with antibiotics"), will be identical to the clinically relevant outcomes of the encounter (the physician examines, agrees that the diagnosis is most likely pneumonia and prescribes antibiotics). Sometimes the patient's judgments and (requests for) actions, differ from the outcomes of the encounter. A concrete example of this from our pool of encounters is the 48 year old teacher with heavy menstrual bleeding, who comes to the gynecologist, starting the encounter by saying: "I want my uterus removed". She makes a strong case that this action "will give me my life back". The physician thinks removing the uterus is too drastic: "why shoot a fly with a canon?" he says. He takes her preferences seriously, informs and examines her thoroughly before he advocates for an ablatio plus an intrauterine device. The patient takes convincing. But when the consultation is about to end, she says: "You have managed to turn me around". The physician responds: "Even a teacher at that...". They both laugh. The patient says: "If you knew how determined I was...". The physician says: "You don't know how happy I would be if you don't have to remove your uterus". The patient concludes: "Let's try that (ablatio plus IUD) then".

Protected by copyright, including for uses related to text and data mining, Al training, and similar technologies.

This encounter is described in detail in another paper in Social Science & Medicine by a linguist working with our group.

(http://www.sciencedirect.com/science/article/pii/S0277953615302434?via%3Dihub).

We find these dialogs extremely interesting, and just as Reviewer 2, we are studying medical decision-making in order to understand why there is not more patient involvement in decisions, in order to increase both physician and patient awareness and hopefully contribute to a culture shift where patient involvement will be the norm rather than the exception.

But for now, the norm is that a clinically relevant action or judgment is (almost exclusively) made and conveyed by the physician. The aim of this paper is to describe the frequency and distribution of these statements. Our study sheds light to the vast amount of decisional work that is done within medical encounters, and is a stepping stone to increased knowledge and awareness about when patients could/should be involved in the decision-making process.

Reviewer 2: In reading the revision, I still have the same major concerns as with the initial paper. Given that they are not presenting this taxonomy here, it is not clear how to handle this fundamental disagreement. It may rest on the editor to decide whether the paper, using the taxonomy as published previously, has a net benefit for readers.

Response to Reviewer 2: As researchers our ambition has been to explore enhance our understanding of complex phenomena. We accept that there are different opinions about our work and that some may prefer a narrow definition of clinical decisions, and we are well aware of those who target and study single and predefined decisions related to a specified topic. What we have done is to develop a novel taxonomy, based on empirical research, and we used this taxonomy to study clinical encounters. We think our work is both transparent and provides new and valuable insights to the field of medical decision-making.